

University of Nebraska - Lincoln

DigitalCommons@University of Nebraska - Lincoln

Faculty Publications in Computer & Electronics Engineering (to 2015) Electrical & Computer Engineering, Department of

2011

Dynamic Video Object Detection with Single PTU Camera

Yun Ye

University of Nebraska-Lincoln, yye@huskers.unl.edu

Song Ci

Institute of Acoustics, sci2@unl.edu

Yanwei Liu

Institute of Acoustics

Hui Tang

Institute of Acoustics

Follow this and additional works at: <https://digitalcommons.unl.edu/computerelectronicfacpub>



Part of the [Computer Engineering Commons](#)

Ye, Yun; Ci, Song; Liu, Yanwei; and Tang, Hui, "Dynamic Video Object Detection with Single PTU Camera" (2011). *Faculty Publications in Computer & Electronics Engineering (to 2015)*. 98.
<https://digitalcommons.unl.edu/computerelectronicfacpub/98>

This Article is brought to you for free and open access by the Electrical & Computer Engineering, Department of at DigitalCommons@University of Nebraska - Lincoln. It has been accepted for inclusion in Faculty Publications in Computer & Electronics Engineering (to 2015) by an authorized administrator of DigitalCommons@University of Nebraska - Lincoln.

Dynamic Video Object Detection with Single PTU Camera

Yun Ye*, Song Ci^{*†}, Yanwei Liu[†], Hui Tang[†]

*Department of Computer and Electronics Engineering, University of Nebraska-Lincoln, USA

[†] Institute of Acoustics, Chinese Academy of Sciences, China

Email: yye@huskers.unl.edu

Abstract—This paper presents a video object detection method under dynamic background with single PTU (pan-tilt unit) camera. The overall procedure contains two steps. First the moving object is tracked through interactive Mean Shift estimation and PTU control. Camera angular speed is updated online according to the estimated object position, whereby the object can remain in the center area of the image plane. In the second step, the foreground is detected through background subtraction and shape contour. An adaptive correlation thresholding method is applied to mitigate the detection distortion in boundary areas. Variational level set contour is then applied to further remove noises and locate moving areas. Our contribution is to incorporate in this procedure the special property of camera movement on a PTU to realize automatic tracking and efficient background matching for motion detection. As demonstrated in the experiment, the proposed method well outlines the foreground objects.

I. INTRODUCTION

Object detection is a primary task in video surveillance applications. An automatic, intelligent surveillance system should be able to monitor moving objects within certain areas and extract the most important information. Therefore both tracking and detection are critical to the performance of such a system. Furthermore, the method adopted in each system must take into account the specific camera arrangement. For example, if multiple cameras are placed at different locations with static pose, the tracking process should be able to coordinate these camera outputs so that the detection process knows how to locate useful information. If the system has control over camera movement, usually fewer cameras are required, with increased complexity in software algorithms since the tracking process needs to incorporate camera control and detection has to deal with changing background.

Object detection with fixed camera often takes advantage of static background. A commonly used technique is background subtraction based on Gaussian Mixture Model (GMM) [1]. This temporal learning process models different conditions of a pixel at certain position as a mixture of Gaussian distributions. Weight, mean and variance values of each Gaussian model can be updated online, and pixels not conforming to any background model are quickly detected. The adaptive learning property makes this technique suitable for real-time applications, and a variety of detection methods are developed combining other spatiotemporal processing [2]–[4]. For object detection with moving cameras, more factors need to be considered. A freely moving camera is usually infeasible in an unmanned surveillance environment due to technical or fiscal constraints. A more practical scheme is to use a PTU camera, where the camera projection center is generally unchanged and the retinal plane is capable of angular movement. In this kind of system, the camera control algorithm for tracking process needs to estimate the angular speed/acceleration of the moving object, and background alignment in different video frames is required for motion detection. In the PTU camera tracking algorithm proposed by Petrov *et al.* [5], a linear feedback controller is applied based on the theory of Lyapunov Stability. The control parameters are updated by object position estimated using Mean Shift [6] algorithm. Mean Shift is efficient in locating object position according to the object's color distribution [7], [8]. However, it can only indicate the approximate position, rather than the shape of the tracked object. Hence further processing is necessary to bring out more accurate information. Among all the features that are useful for shape detection, motion is one of the most essential. The first step for motion detection with a moving camera is background alignment. While traditional optical flow approach is very time-consuming

and loses accuracy when the camera has large motions [7], features such as corners and edges [9], [10] are often used to estimate the geometric relation between different video frames. In an affine camera model, this relation can be represented by either the fundamental matrix [11] or the homography [10]. We will show in the next section that the fundamental matrix is not suitable for a PTU tracking system, and the multi-layer homography model [12] is unnecessary, because of the special property of the camera movement.

In this paper, we propose an object detection method based on the analysis of PTU camera model. In the tracking process, one PTU camera is used. The linear control algorithm in [5] is used to follow the object's movement. The Mean Shift estimation and the control algorithm work in an interactive way to enhance accuracy and speed up convergence. The moving object can thus always be visible in center area of the image plane. In the detection process, first background is aligned using the homography computed from RANSAC fitting. Then correlation between the video frame and aligned background image is adaptively thresholded according to the distance between each pixel and the estimated object center from Mean Shift. A variational level set method [13] is applied afterwards to contour the outline and remove noises. Experiment on our PTU camera video shows promising results.

II. PTU CAMERA MODEL

This section describes the unique geometric property of a PTU camera model, i.e. the camera projection center remains unchanged while the pan and tilt angles are controllable. The tracking and detection process in our proposed method are designed based on this property.

A. Object tracking

Figure 1 (a) illustrates the PTU camera control algorithm proposed in [5]. Focus F denotes the projection center. The image plane is viewed down along its y axis, and is projected on the X - Y world coordinate plane. α is the angle between the object center and X axis, θ is the angle between the image center and X axis, f is the focal length, and x_c is the distance between projected object center and image center along x axis. Only pan control is displayed in the figure. The algorithm applies to tilt control similarly.

In the linear feedback controller, the Lyapunov function aims to minimize x_c and the difference between estimated object speed and measured object speed. The

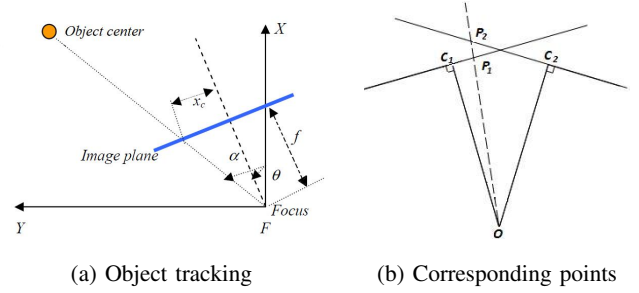


Fig. 1. PTU camera model.

camera angular speed $\omega_{\theta k}$, camera angle θ , and estimated distance x_c are updated at every time instance,

$$\omega_{\theta k} = \hat{\omega}_{\alpha k} + p x_{ck} \quad (1)$$

$$\theta_{k+1} = \theta_k + \omega_{\theta k} \Delta t_k \quad (2)$$

$$\hat{x}_{ck+1} = f \tan(\hat{\alpha}_{k+1} - \theta_{k+1}) \quad (3)$$

where $\hat{\omega}_{\alpha}$ is the estimated object angular speed, and $\hat{\alpha}$ is the estimated object angle. p is a parameter used to control the convergence speed in the dynamic system. Δt_k is the duration of the k -th time interval for control parameters update. Different from [5] where constant time interval is applied, we consider the control delay in the real communication system and adopt noisy interval values in equation (2) to test the performance of our proposed method.

B. Homography

With a moving camera, object detection has to consider the background difference among the video sequence. Background alignment for motion detection is usually implemented by computing the fundamental matrix [11] or homography [10]. However, the fundamental matrix method fails to handle the situation when the camera projection center is not moving. In this case, the computed matrix is singular and it is very sensitive to noises. By contrast, the homography method estimates the direct mapping between corresponding points. And there is no need to worry about the multi-layer homography problem described in [12] since a point on one projection ray will always be projected to a fixed position in the image plane regardless of the depth of this point. As demonstrated in Figure 1 (b), if a point locates on the ray OP_1 , where O is the projection center, and P_1 is its projected point in image plane C_1 , its projection on image plane C_2 will be P_1 , no matter how far it is from the projection center.

The background alignment results with these two methods are shown in Figure 2. The alignment process with corner point detection and matching will be introduced in following section. As seen in the ceiling light area, the mismatched pairs in Figure 2 (b) due to similar correlation values are eliminated in Figure 2 (c) by the homography method, and the background is correctly aligned with the video frame.

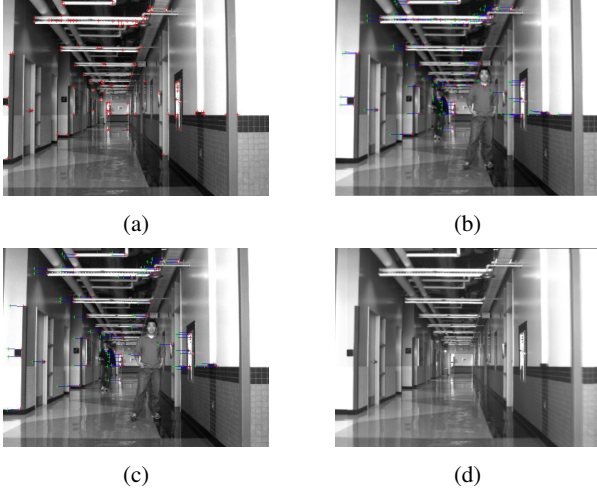


Fig. 2. (a) Background image with detected corners; corner pair matching by (b) fundamental matrix method and (c) homography method; (d) aligned background.

III. SYSTEM ARCHITECTURE

The system architecture of the proposed two step object detection method is demonstrated in Figure 3.

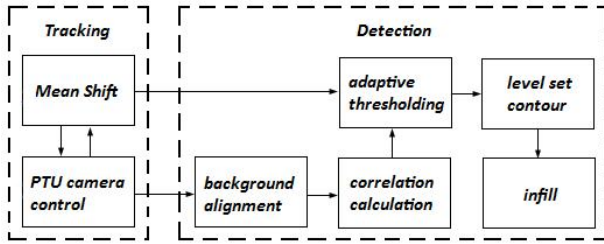


Fig. 3. The system structure of the proposed method.

In the tracking process, two modules work interactively. The Mean Shift algorithm provides the measured object position as input for parameters update in PTU camera control, while the updated control parameters provide initial position for Mean Shift at next time instance in order to speed up the convergence. Specifically, the estimated distance \hat{x}_c from equation (3) is provided as initial position for the Mean Shift algorithm

to speed up the convergence. And in turn the output x_c from Mean Shift is used in equation (1) to update the camera angular speed at next time instance. In the presence of occlusion, the camera will stop moving while the calculated Bhattacharyya coefficient is below certain threshold.

In the detection process, firstly background is aligned with each video frame. A background image necessary for subtraction can be synthesized either from several pre-captured backgrounds [10], or from some learning methods [14], [15] using other video frames. In the alignment process, Harris corner finder [9] is applied to detect feature points in both video frame and the background image. Corners in one image find their matches in the other with highest correlation. And mismatched pairs are eliminated by RANSAC fitting, through which the homography is computed:

$$\min(\|Hx' - x\|^2) \quad (4)$$

Here H is the homography, and x' , x denote the coordinates of a matched pair in two images. Secondly, the correlation value of each pixel in video frame with aligned background is calculated and thresholded according to its distance from the object center (x_c, y_c) estimated by Mean Shift algorithm,

$$T(x, y) = A \exp\left(-\frac{(x - x_c)^2 + (y - y_c)^2}{B^2}\right) \quad (5)$$

$$I'(x, y) = \begin{cases} I(x, y), & \text{if } \text{Corr}(x, y) < T(x, y) \\ 0, & \text{otherwise} \end{cases} \quad (6)$$

The parameter A and B are predefined constants used to adjust the range of the threshold T . By adopting such an adaptive measure, foreground areas around the object center will have higher threshold and can be more easily detected than boundary areas, where the lighting condition is volatile and the subtraction result is prone to suffer.

After thresholding, the variational level set method proposed by Li *et al.* [13] is applied on the residual image to contour and infill the outline of foreground objects. Noises can be further removed due to the length shrinking property of the evolutionary curves. During the implementation of the detection process, multi-resolution strategy is adopted to reduce computation.

IV. EXPERIMENTAL RESULTS

The experimental results with our PTU (Directed Perception D47) camera tracking system are demonstrated

in Figure 4. Two different video frames are displayed in Figure 4 (b) and (c) to illustrate the tracking result. The walking person in red (courtesy of Dalei Wu) is always visible in center area. Background alignment can be observed in Figure 2. Correlation thresholding results with constant threshold values are provided for comparison. The threshold value is set to 0.3 (Figure 4 (d)) and 0.7 (Figure 4 (e)) respectively. The constants defined in equation (5) are selected such that $A = 0.8$, and B equals to one fourth of the image diagonal length. As shown in Figure 4 (f) - (g), our adaptive thresholding method is more robust to noises, thus can facilitate the level set contour process. As can be observed from the outline infill result in Figure 4 (h), the proposed detection method contours moving object shapes efficiently, even only intensity values of the video data are processed.

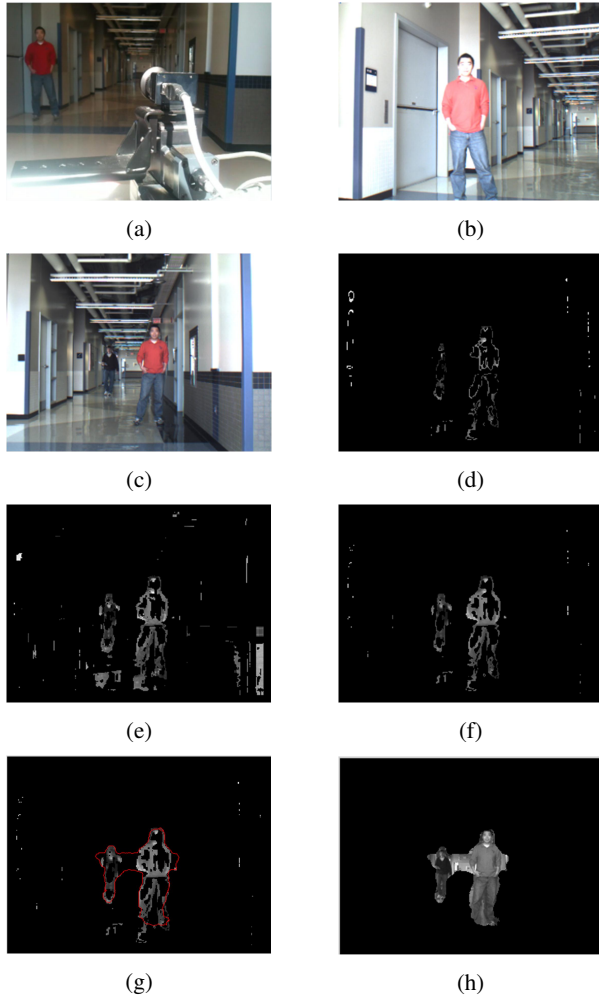


Fig. 4. (a) PTU camera; tracking (b) - (c) and detection (d) - (h). Correlation thresholding with (d) - (e) constant and (f) adaptive threshold; (g) level set contour; (h) infill.

V. CONCLUSIONS

The video object detection method presented in this paper utilizes the special geometric property of PTU camera movement. In the tracking process, Mean Shift and the camera control model work interactively to enhance accuracy and achieve real-time performance. This tracking mechanism centers the object in image plane and provides useful information for detection process. In the detection process, foreground areas are efficiently detected through background alignment and correlation thresholding. The adaptive thresholding scheme works successfully given the tracking position information. Experimental results on intensity data are encouraging. The proposed method is suitable for resource limited surveillance applications.

REFERENCES

- [1] C. Stauffer and W. Grimson, "Learning patterns of activity using real-time tracking," *PAMI*, vol.22, no.8, pp.747-757, Aug.2000.
- [2] S. Babacan and T. Pappas, "Spatiotemporal algorithm for background subtraction," *ICASSP 2007*, vol. 1, pp. I-1065 -I-1068.
- [3] P. Suo and Y. Wang, "An improved adaptive background modeling algorithm based on gaussian mixture model," *ICSP 2008*, 2008, pp. 1436 -1439.
- [4] J. Gallego, M. Pardas, and G. Haro, "Bayesian foreground segmentation and tracking using pixel-wise background model and region based foreground model," *ICIP 2009*, pp.3205-3208.
- [5] O. B. Plamen Petrov and K. Muratovski, "Face detection and tracking with an active camera," *ICIS 2008*, pp. 14-34 -14-39.
- [6] D. Comaniciu and P. Meer, "Mean shift: a robust approach toward feature space analysis," *PAMI*, vol. 24, no. 5, pp. 603 -619, May 2002.
- [7] J.-W. Hsieh and J.-X. Lee, "Video object segmentation using kernel-based models and spatiotemporal similarity," *ICIP 2006*, pp. 1821 -1824.
- [8] F. Shimin, G. Qing, X. Sheng, and T. Fang, "Human tracking based on mean shift and kalman filter," *AICI 2009*, vol. 3, nov. 2009, pp. 518 -522.
- [9] C. Harris and M. Stephens, "A combined corner and edge detector," *Proceedings of the 4th Alvey Vision Conference*, 1988, pp. 147-151.
- [10] J.-W. Hsieh, "Fast stitching algorithm for moving object detection and mosaic construction," *ICME 2003*, vol.2, pp.85-88.
- [11] P. D. Kovesi, "MATLAB and Octave functions for computer vision and image processing," Centre for Exploration Targeting, School of Earth and Environment, The University of Western Australia, available from: <<http://www.csse.uwa.edu.au/~pk/research/matlabfns/>>.
- [12] Y. Jin, L. Tao, H. Di, N. Rao, and G. Xu, "Background modeling from a free-moving camera by multi-layer homography algorithm," *ICIP 2008*, pp. 1572 -1575.
- [13] C. Li, C. Xu, C. Gui, and M. Fox, "Level set evolution without re-initialization: a new variational formulation," *CVPR 2005*, vol. 1, pp. 430 - 436 vol. 1.
- [14] Z. Kim, "Real time object tracking based on dynamic feature grouping with background subtraction," *CVPR 2008*, pp. 1 -8.
- [15] Y. Sheikh, O. Javed, and T. Kanade, "Background subtraction for freely moving cameras," *ICCV 2009*, pp. 1219 -1225.